Translating Cancer Complexity to Clinical Decisions

David P. Carbone, MD, PhD

Cells are complex machines that sometimes get out of control and take over their host, i.e., become cancer cells. It goes without saying that to control or repair them, a thorough understanding of how they work and a specific understanding of what has gone awry is required. The study by Chen et al.,[1] led by one of the pioneers in this field, is a technically rigorous step toward sorting these errant machines into broad categories that may require different therapeutic approaches. In this study, Chen et al. defined a predictor of survival based on data from 680 tumors and defined a clinically feasible reverse transcriptase polymerase chain reaction classifier that significantly classified a test set of tumors for survival. It is thus a significant study for the size of the training set, the rigor of the analysis, and the translation to a reverse transcriptase polymerase chain reaction assay in a 101-patient test set. Such classification attempts, and in particular attempting to understand what these classifications imply for patient management, are clearly the next frontier in oncology. The last sentence of the abstract looks toward clinical application and states that these data should be incorporated "into new clinical trials with the goal of personalized treatment of lung cancer patients and improving patient survival." For example, the discovery and definition of subsets of these tumors driven by epidermal growth factor receptor (EGFR) mutations and ALK fusions have revolutionized both our understanding of how cancers work and our therapies for these subsets.

Definition of clinically actionable subsets by expression array technology, as described in the article, has to date been much less successful, at least in part because of two significant issues: (1) they are typically tied to a particular way of looking at the cancer cell, ignoring all others, and (2) they fail to carefully define the clinical problem they are addressing and choose sample sets and analyses appropriate to answer this question rather than whatever is conveniently available. This study, although far from unique in this respect, suffers from both these issues.

Regarding the first problem, it is now clear that cancer cells are not completely defined and cannot be completely understood by studying their pattern of acquired genetic mutations or by studying RNA expression patterns in isolation. EGFR-mutant tumors are clearly different from EGFR wild-type tumors in almost every way—patients with these tumors on average live longer no matter how you treat them, respond better to chemotherapy, and have different patterns of relapse and metastasis. There are a handful of mutant genes known today (perhaps about 10) that logically define significant subsets of lung cancers, and future studies attempting to define RNA expression classifiers should include these gene mutation variables in their analyses. Important RNA patterns associated with a clinical phenotype might be completely different in each of these gene mutation subsets, and by mixing these all together in the training set, we unnecessarily muddle our data. Although "multidimensional" classification analytic techniques need to be optimized, it is clear that future classification studies should take into account at least these dominant subsets. This study does not do this, even lumping adenocarcinomas with squamous carcinomas, each with clear differences in gene mutation patterns. Conversely, it is also clear that all tumors with EGFR mutations, and even all tumors with exactly the

Vanderbilt University Medical Center, Nashville, Tennessee.
Disclosure: The author declares no conflicts of interest.
Address for correspondence: David P. Carbone, MD, PhD, Vanderbilt University Medical Center, 685 PRB, 2220 Pierce Avenue, Nashville 37232-6838, Tennessee. E-mail: d.carbone@vanderbilt.edu

ISSN: 1556-0864/11/0609-1455

same EGFR mutation, do not behave alike, and thus RNA expression classifiers may help define different behaviors within these genetic subsets.

The second problem is as important and perhaps more subtle but just as pervasive. To answer a question, the question must be clearly defined, and one must use well-defined and appropriate datasets to find useful answers. The goal of this study is to define a "prognostic" classifier (often defined as outcome independent of therapy) and not a "predictive" one (defining which patients will benefit from therapy). However, it is very important to define the projected use of such a "prognostic" classifier. It is stated in the Introduction of the article that such risk classifiers are useful in defining patients destined to relapse and thus might benefit from adjuvant therapy. Adjuvant chemotherapy clearly improves survival, and adjuvant radiation clearly alters the risk of local relapse. Thus, ideally, clinically useful prognostic classifications should assist in the definition of a subset of patients who are at high risk for relapse without adjuvant therapy, and within this subset, there are hopefully tumors, defined by a predictive classifier, that become low risk when treated with adjuvant chemotherapy or radiation.

Therefore, for this purpose, this prognostic classifier should have been derived only using patients who received no adjuvant therapy. The inclusion of patients who received adjuvant therapy into the training set dooms the results to failure in this regard. Specifically, individual patients at high risk of relapse without adjuvant therapy, but who benefited from adjuvant therapy and became low risk (did not relapse) because of it, would be falsely forced into the low-risk group of patients by this approach. Resulting classifiers would only identify patients who do poorly and don't do any better with adjuvant therapy. The subset analysis presented in the article showed that after definition of this classifier, high-risk patients showed no benefit from adjuvant chemotherapy, which confirms this problem. If the question is definition of mechanisms driving relapse, active pathways in untreated patients who relapse should be compared with untreated patients who don't, and this may define novel therapy targets that may alter this behavior.

The use of patients homogeneous for the relevant clinical questions is thus absolutely essential. Many of the patients in this study had "unknown" adjuvant treatments. Samples with incomplete clinical annotation should not be included at all in any modern analysis, and treated patients should not be lumped with untreated ones, if the intent is to identify patients at risk for relapse and who might benefit from these therapies. The era of definition of classifiers using "samples of convenience" accumulated for the purpose of large numbers and good p values should end.

Although difficult, accounting for the major genetic subsets in expression array studies, defining the exact question being asked, and using appropriate clinical cohorts should be expected today. As our state of knowledge progresses, the general approach to defining broad categories of tumors that behave more or less alike with only moderate degrees of certainty will hopefully give way to a day when tumors are characterized by possibly dozens of potently predictive patterns of inherited differences, acquired mutations, altered gene and protein expression, post-translational modifications, and signaling patterns, with therapies designed to match each pattern. We need to stop treating lung cancer as a single disease at all levels of analysis and clearly define our clinical goals when designing our research experiments.

REFERENCE

1. Chen G, Kim S, Taylor JMG, et al. Development and validation of a qRT-PCR classifier for lung cancer prognosis. *J Thorac Oncol* 2011;6: 1481–1487.

Response:

Dr. David Carbone makes several points on our manuscript which he feels are problematic and suggests clinically-actionable subsets by expression array technology, as described in our paper are not successful because: 1) they are typically tied to a particular way of looking at the cancer cell, ignoring all others, and 2) they fail to carefully define the clinical problem they are addressing and choose sample sets and analyses appropriate to answer this question rather than whatever is conveniently available. Although we do not agree with his conclusions, we think they are interesting and important issues and we welcome the opportunity to discuss them. As detailed in our manuscript by Chen et al, we hypothesized that there are diverse survival-related processes that are associated with lung adenocarcinomas and we then actively attempted to incorporate this heterogeneity into a model by using representation of small numbers of genes from as many diverse survival-related clusters (presumably cell processes) as possible. This is in contrast to many models where the most significant survival-related genes are used regardless of whether they are associated with the same underlying biological process or not. Our goal was to develop a model that was broadly useable for a diverse tumor population such as that seen in lung adenocarcinomas and where we as yet do not know all of the processes that are critical for defining patient outcome. While we do not disagree with his comments regarding ALK fusions or EGFR-mutant lung adenocarcinomas, these tumors which are from non-smokers are clearly genomically more stable, and quite unlike the vast majority of lung cancers. In fact, Dr. Carbone admits that there are a handful of genes that may significantly define subsets; we humbly submit that these form a small slice of the pie which are defined by complex alterations in multiple pathways. Although it is tempting to hope that gene mutations will clearly define other subgroups as Dr. Carbone suggests, analyses done with our collaborators (Ding et al, Nature, 2008) reveal that lung adenocarcinomas often have very large numbers of simultaneous mutations, many occurring at low frequency and unfortunately may not neatly divide